

**STATISTICAL MODELING FOR ROAD CROSSING
BEHAVIOR OF PEDESTRIANS ON URBAN ROADS**

MARIYA KHATOON



**DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY DELHI
AUGUST 2015**

© Indian Institute of Technology Delhi (IITD), New Delhi, 2015

STATISTICAL MODELING OF ROAD CROSSING BEHAVIOR OF PEDESTRIANS ON URBAN ROADS

by

MARIYA KHATOON

Department of Mathematics

Submitted

in fulfillment of the requirements of the degree of

Doctor of Philosophy

to the



Indian Institute of Technology Delhi

August 2015

Dedicated to my family

Certificate

It is hereby certified that the thesis entitled “**Statistical modeling of pedestrian behavior on urban roads**” submitted by **Ms. Mariya Khatoon** to the Indian Institute of Technology Delhi, for the award of the degree of **Doctor of Philosophy**, is a record of the original bonafide research work carried out by her under our supervision and guidance. The thesis work, in our opinion, has reached the requisite standard fulfilling the requirements for the degree of Doctor of Philosophy. The results contained in this thesis have not been submitted in part or in full, to any other university or institute for the award of any degree or diploma.

Dr. Geetam Tiwari

Professor

Department of Civil Engineering

Indian Institute of Technology Delhi

Dr. Niladri Chatterjee

Professor

Department of Mathematics

Indian Institute of Technology Delhi

Date:

New Delhi

Acknowledgments

I am taking this opportunity to thank individually, all those who have been associated in one way or the other, throughout my Ph.D work. First of all, I thank the Almighty Allah for his blessings on me which made me reach this place and always providing me the best of everything.

My first and foremost heart-felt gratitude goes to my supervisors Prof. Niladri Chatterjee and Prof. Geetam Tiwari. I am very much thankful to them for their care and interests on my work, for stimulating discussions, and for their valuable advices during my research work. I thank Prof Niladri Chatterjee for enhancing my writing and presentation skills of scientific document. Apart from supervision of my thesis work, I would like to thank Prof. Geetam Tiwari –my employer and mentor. From the past six years of working and learning with her has been a journey of consistent professional and personal growth.

I acknowledge TRIPP (Transport Research and Injury Prevention Programme)/IITD, India for providing me financial support, lab facilities, assistant in attending the national and international conferences and interaction with the experts in field during the research period. I am also grateful to Prof. Dinesh Mohan from TRIPP for the various insights into the field studied in this research.

I also thank IIT Delhi authorities for providing the essential academic as well as campus facilities for pursuing the research.

I express my regards and Special thanks to my SRC (Student Research Committee) members: Prof. R. K. Sharma (Dept. of Mathematics), Dr. Aparna Mehra (Dept. of Mathematics), and Dr. R.R. Kalaga (Dept. of Civil Engineering) for their valuable comments and suggestions at different stages of the research.

I am grateful to my pre-Ph.D course instructors: Dr. Wagish Shukla, Prof. Dinesh Mohan, Prof Geetam Tiwari, Prof. R. Chattopadhyay and Dr. Dipayan Das.

The colleagues and staff at TRIPP have been very supportive of my work and commitments during all this time. I am quite fortunate to have colleagues like Himani Jain, Deepty Jain, Shalini Rankavat, Rahul Goel, Sumit Gupta and Col. SSLN Sarma with whom I had spend wonderful time and earned a lot of cherish-able moments. Thanks for all the care, love, understanding, and encouragement.

I express my heartfelt gratitude to (Late) Mr Ashok Sharma who encouraged me to do the Ph.D. My parents, parents-in-law, sisters and brother have been patient, supportive and proud of my academic and professional achievements, and I owe my self-confidence to them. My heartfelt thanks to my mother Shakeela Mirza who has been an inspiration and an anchor during the entire research phase and my life,

I dedicate this 5 years worth of effort to my son Hasan Khan and my husband Yusuf Khan. Hasan has adjusted with different caretakers without any complain during the research period. I thankfully wish him a very bright future. And last but not the least, heartfelt thanks to my husband for his love, affection, support, and the unfailing faith he has on me.

Mariya Khatoon

Abstract

This thesis explored the utility of binary probabilistic modeling and its performance measures to analyze the pedestrians' unsafe road crossing behavior exhibited by different types of pedestrians and their associated risk at different traffic and road environments.

For the analysis we have collected the data at six different types of pedestrian crosswalks in Delhi. The selected locations are signalized crosswalks and non-signalized crosswalks such as near bus stops, foot of flyovers/grade separators, and free left turn.

In order to achieve higher prediction accuracy, utility of both most used binary Probit and Logistic regression have been explored. Binary Probit and Logistic regression models provide approximately similar results. However, Logistic regression model is marginally better in depicting the pedestrian road crossing behavior. Therefore, the detailed analysis of the data has been done with Logistic regression model. Pedestrians cross the road unsafely with the gaps available in the traffic. Therefore, gap size parameter is found highly significant variable to determine the pedestrian road crossing behavior.

The adequacy of the fitted models has been checked by analyzing the residuals and influential observations. No significant inadequacy has been found in the models. For the models evaluation of diagnostic statistics and their graphical display detected some cases as outliers. But, it is observed that they did not contribute significantly in parameter estimation and summary measures of model fits.

At the end, we have done the performance analysis i.e. validation of the developed Logistic regression models. Apparent validation has been used to measure how effectively a model

describes the outcome variable in the sample data set. Furthermore, internal validation techniques have been used to measure the performance of the model in the population from which the sample has been taken and these techniques have been taken as an approximation to external validation.

The results of the study are the basic inputs to understand the road crossing behavior of pedestrian and their associated risk needed to design well structured transport system.

Contents

Certificate.....	i
Acknowledgments.....	iii
Abstract.....	v
List of Figures.....	xiii
List of Tables	xv
1 Introduction.....	1
1.1 Aim of the research	3
1.1.1 Scope of the work	3
1.2 Need for study	3
1.3 Research motivations	4
1.4 Issues in pedestrian behavior and risk modeling.....	5
1.5 Research objectives, hypothesis, and questions.....	5
1.5.1 Objectives	5
1.5.2 Hypothesis.....	6
1.5.3 Research questions.....	6
1.6 Definitions.....	7
1.6.1 Pedestrians' unsafe crossing behavior	7
1.6.2 Pedestrian risk.....	8
1.6.3 Adequate gap size for pedestrian crossing.....	9
1.7 Structure of the thesis.....	10
1.8 Brief methodology of the research.....	14
1.9 Software used	14
2 Literature review	15
2.1 Pedestrian road crossing behavior.....	15
2.1.1 Before and after studies	15
2.1.2 Signalized and non-signalized crosswalks: Road and traffic environment	17
2.2 Pedestrian risk	23
2.2.1 Signalized crosswalks	23
2.2.2 Near bus stop.....	25
2.2.3 Free left turn.....	26

2.3	Review of methods to collect pedestrian data.....	26
2.4	Conclusions	30
3	Data collection and preliminary data analysis	33
3.1	Site selection	33
3.1.1	Impact of change in infrastructure: Before and after the grade separator.....	34
3.1.2	Related to infrastructure type: Signalized, non-signalized, free left turn, and foot of flyover crosswalks	36
3.2	Data extraction	38
3.3	Mean and variance of the data set collected at two foot of flyovers.....	39
3.4	Pedestrians' crossing behavior at different crosswalks.....	40
3.5	Sample size at different crosswalk	42
3.6	Predictor variables.....	44
3.7	Preliminary data investigation.....	50
3.7.1	Change in infrastructure: Before and after the construction of grade separator	50
3.7.1.1	Pedestrian characteristics.....	50
3.7.1.2	Gap analysis.....	51
3.7.1.3	Pedestrian road crossing with different type of conflicting vehicle	54
3.7.1.4	Vehicles' speed analysis.....	55
3.7.1.5	Pedestrians' waiting time analysis.....	56
3.7.1.6	Correlation between waiting time and gaps accepted by pedestrians	56
3.7.2	Signalized and non-signalized crosswalk	58
3.7.2.1	Pedestrians' and conflicting vehicles' characteristics	58
3.7.2.2	Gap size and waiting time analysis.....	59
3.7.2.3	Analysis of rejected and accepted gap sizes according to the volume of traffic	60
3.7.3	Free left turn.....	61
3.7.3.1	Pedestrians' and conflicting vehicles' characteristics	61
3.7.3.2	Gap analysis.....	61
3.7.3.3	Waiting time analysis	62
3.7.4	Foot of flyover	63
3.7.4.1	Pedestrians' and conflicting vehicles' characteristics	63

3.7.4.2	Pedestrians' waiting time analysis.....	63
3.8	Conclusions	65
4	Binary choice regression modeling.....	67
4.1	Comparison measures	68
4.2	Binary choice model: Random utility approach.....	68
4.3	Parameter estimation: Maximum Likelihood Estimation	73
4.4	Binary Probit and Logistic regression models: Transformations of Z_{in}	74
4.4.1	Binary Probit regression	75
4.4.2	Binary Logistic regression	76
4.5	Basic Assumption and requirements of binary Probit and Logistic regression model ..	76
4.6	Assessing the models: Log-likelihood statistics, χ^2 test and Nagelkerke's R_N^2	77
4.6.1	Log-likelihood statistic and χ^2 test.....	78
4.6.2	Nagelkerke's R_N^2	79
4.7	Assessing the contribution of predictor variables: the Wald statistics.....	80
4.7.1	Continuous variables.....	80
4.8	Categorical variables	80
4.9	Binary Choice Modeling: Probit and Logistic	82
4.9.1	Goodness-of-fit tests	83
4.9.2	Wald statistics	84
4.9.3	Probabilities predicted by the Probit and Logistic models	85
4.10	Probit vs. Logistic.....	86
4.11	Conclusions	87
5	Application of binary Logistic regression models.....	89
5.1	Change in infrastructure: Before and after the construction of the grade separator	90
5.1.1	Pedestrians' unsafe crossing behavior	90
5.1.2	Estimation of pedestrian risk	93
5.2	Related to infrastructure type: Signalized, non-signalized, free left turn, and foot of flyover crosswalks	96
5.2.1	Signalized and non-signalized crosswalk near bus stop	100
5.2.1.1	Pedestrians' unsafe crossing behavior	100
5.2.1.2	Probability of road crossing at signalized and non-signalized crosswalks....	103

5.2.1.3	Estimation of pedestrian risk	104
5.2.2	Free left turn.....	105
5.2.2.1	Pedestrians' unsafe crossing behavior	105
5.2.2.2	Estimation of pedestrian risk	106
5.2.3	Foot of flyover	107
5.2.3.1	Pedestrians' unsafe crossing behavior	107
5.2.3.2	Estimation of pedestrian risk	109
5.3	Conclusions	110
6	Analysis of residuals and influential observations	111
6.1	Residuals in Binary Logistic model	114
6.2	Explanatory residual Analysis.....	118
6.2.1	Descriptive statistics of residuals.....	118
6.2.2	Residual plots.....	120
6.3	Influential observations	131
6.3.1	Leverage.....	131
6.3.2	Cook's distance.....	132
6.3.3	DFBeta value	133
6.3.4	Descriptive statistics for the influence of the observations in the model	134
6.3.5	Plots for the influential statistics.....	135
6.3.6	Influence of an outlier.....	139
6.4	Conclusions	142
7	Validation of binary Logistic regression model	145
7.1	Apparent validation	147
7.1.1	Nagelkerke's R_N^2	147
7.1.2	Log-likelihood statistic and χ^2 test.....	148
7.1.3	Hosmer–Lemeshow test.....	149
7.1.4	Correct Classification Rate (CCR).....	152
7.1.4.1	Sensitivity and Specificity	153
7.1.4.2	Classification plot.....	154
7.1.4.3	Receiver Operating Characteristic (ROC) curves	157
7.2	Internal validation	159

7.2.1	Data-splitting.....	159
7.2.2	Cross-validation	162
7.2.2.1	Types of cross-validation.....	162
7.2.3	Bootstrapping.....	167
7.3	External validation	172
7.4	Conclusion.....	172
8	Conclusions, limitations of study, and future scope of research.....	175
8.1	Conclusions of the study	177
8.1.1	Data collection and preliminary data analysis	177
8.1.2	Binary choice regression modeling.....	178
8.1.3	Applications of Binary Logistic regression models.....	180
8.1.4	Analysis of residuals and influential observations.....	183
8.1.5	Validation of binary Logistic regression model.....	185
8.2	Research contributions	188
8.3	Limitations of study and future scope of the work.....	190
	Bibliography	195
	Annexure.....	215
	Author's Bio-data.....	225

List of Figures

Figure 1-1: Methodology of the research.....	14
Figure 3-1: Schematic intersection drawing before the construction of the grade separator.....	35
Figure 3-2: Schematic intersection drawing after the construction of the grade separator	36
Figure 3-3: Schematic drawing of IIT Delhi intersection	38
Figure 3-4: Example illustrating the definition of gap size parameter	47
Figure 5-1: Plot of Pedestrians' road crossing probability with the different gap sizes before the grade separator	95
Figure 5-2: Plot of Pedestrians' road crossing probability with the different gap sizes after the grade separator	95
Figure 6-1: Residual plot between the ordinary residuals and the predicted probabilities	121
Figure 6-2: Residual plot between the Logit residuals and the predicted probabilities	122
Figure 6-3: Residual plot between the Logit residuals and the predicted probabilities without case 470.....	122
Figure 6-4: Residual plot between the Pearson residuals and the predicted probabilities	123
Figure 6-5: Residual plot between the Pearson residuals and the predicted probabilities without case 470.....	124
Figure 6-6: Residual plot between the Deviance residuals and the predicted probabilities	124
Figure 6-7: Residual plot between the Normalized residuals and the predicted probabilities....	125
Figure 6-8: Scatter plot between the Leverage values and the predicted probabilities	135
Figure 6-9: Scatter plot between the values of Cook's statistics and the predicted probabilities	136
Figure 6-10: Scatter plot between the value of DFBeta statistics and the predicted probabilities	138
Figure 7-1: Classification Plot for the model developed on non-signalized crosswalk.....	155
Figure 7-2: ROC curve for the model developed on non-signalized crosswalk	158

List of Tables

Table 1-1: Share of pedestrian fatalities in Delhi from 2001-2012	2
Table 3-1: Mean and variance of rejected and accepted gap of pedestrian at two foot of flyovers	40
Table 3-2: Mean and variance of waiting time of pedestrian at two foot of flyovers.....	40
Table 3-3: List of the selected predictor variables from the past research	45
Table 3-4: Correlation coefficients of the predictor variable traffic volume with the other predictor variables.....	49
Table 3-5: Characteristics of pedestrians and their type of crossings before and after the construction of grade separator	51
Table 3-6: Rejected and accepted gap analysis: Before and after the grade separator	53
Table 3-7: Gap acceptance with respect to conflicting vehicles: Before and after the grade separator	54
Table 3-8: Mean speeds of conflicting vehicles: Before and after the grade separator	55
Table 3-9: Mean waiting time of pedestrian at origin and median: Before and after the grade separator	56
Table 3-10: Characteristics of pedestrians and type of crossings at signalized and non-signalized crosswalk.....	58
Table 3-11: Speed of different types of conflicting vehicles at signalized and non-signalized crosswalk.....	59
Table 3-12: Levene's test and t-values for the gap size and waiting time at signalized and non-signalized crosswalks.....	59
Table 3-13: Count of rejected and accepted gaps at high and low traffic volume.....	60
Table 3-14: Characteristics of pedestrians and type of crossings at the free left turn	61
Table 3-15: Speed of different types of conflicting vehicles at the free left turn	61
Table 3-16: Rejected and accepted gap analysis at the free left turn.....	61
Table 3-17: Pedestrians' waiting time analysis at the free left turn.....	62
Table 3-18: Characteristics of pedestrians and type of crossings at the foot of flyover	63
Table 3-19: Speed of different types of conflicting vehicles at the foot of flyover.....	63
Table 3-20: Results of ANOVA for pedestrians' waiting time at the foot of flyover	64
Table 4-1: Dummy coding for the variable "type of conflicting vehicle"	82
Table 4-2: Goodness-of-fit measures of Probit and Logistic regression model	83
Table 4-3: Wald Statistics of the predictor variable in binary Probit and Logistic regression models	85
Table 4-4: Probabilities of road crossing for 10 cases predicted by Probit and Logistic regression models.....	86

Table 5-1: Results of Logistic regression for the model developed on before and after the grade separator.....	91
Table 5-2: Probability (%) of road crossing at different gap sizes: Before and after the grade separator.....	94
Table 5-3: Results of Logistic regression for the combine model considering type of crosswalk as one of the predictor parameter	98
Table 5-4: Results of Logistic regression for the model developed at signalized and non signalized crosswalks.....	101
Table 5-5: Probability (%) of road crossing with respect to different gap sizes at the signalized and non-signalized crosswalks.....	104
Table 5-6: Results of Logistic regression for the model developed at the free left turn.....	105
Table 5-7: Probability (%) of road crossing with respect to different gap sizes at the free left turn	107
Table 5-8: Results of Logistic regression for the model developed at the foot of flyover	108
Table 5-9: Probability (%) of road crossing with respect to different gap sizes at the foot of flyover	109
Table 6-1: Descriptive statistics of the Residuals form the model developed on the non-signalized crosswalk	119
Table 6-2: Values of the residuals for the unusual observation (case 470) from the model developed on the non-signalized crosswalk.....	125
Table 6-3: Descriptive statistics of the Residuals form the model developed on the before the construction of the grade separator	128
Table 6-4: Descriptive statistics of the Residuals form the model developed on the after the construction of the grade separator	128
Table 6-5: Descriptive statistics of the Residuals form the model developed on the foot of flyover	129
Table 6-6: Descriptive statistics of the Residuals form the model developed on the free left turn	129
Table 6-7: Descriptive statistics of the Residuals form the model developed on the Signalized crosswalk.....	130
Table 6-8: Descriptive statistics of the influential observations for the model developed on the non-signalized crosswalk	134
Table 6-9: Values of the Influential statistics for the outlier observation (case 470) from the model developed on the non-signalized crosswalk.....	139
Table 6-10: Descriptive of the influential statistics for the model developed on the before the construction of the grade separator	140
Table 6-11: Descriptive of the influential statistics for the model developed on after the construction of the grade separator	140
Table 6-12: Descriptive of the influential statistics for the model developed on the foot of flyover	141

Table 6-13: Descriptive of the influential statistics for the model developed on the free left turn	141
Table 6-14: Descriptive of the influential statistics for the model developed on the signalized crosswalk.....	142
Table 7-1: Nagelkerke's R_N^2 , -2 Log-likelihood and χ^2 tests	149
Table 7-2: Hosmer-Lemeshow Test.....	150
Table 7-3: Contingency Table for Hosmer-Lemeshow test at the non-signalized crosswalk	151
Table 7-4: Possible outcomes from a binary classifier	152
Table 7-5: Sensitivity, Specificity and overall correct prediction percentage	153
Table 7-6: Sensitivity, Specificity and overall correct percentage at different cut-off points of the model developed on a non-signalized crosswalk.....	156
Table 7-7: Test results for ROC curve.....	157
Table 7-8: Results of Spilt sample validation for the model developed on non-signalized crosswalk.....	160
Table 7-9: Results of the Spilt sample validation for the fitted models.....	161
Table 7-10: Classification accuracy from 10 fold cross-validation for the model developed on non-signalized crosswalk.....	164
Table 7-11: Results of 10 fold Cross validation for the developed models.....	166
Table 7-12: Bootstrap Specifications.....	168
Table 7-13: Bootstrapping results for the model developed on the non-signalized crosswalk ..	170
Table 7-14: Biases statistic range for the developed models.....	171